2

Amendments to the Claims

1. (Original) A system, comprising:

a psycho-physical state detection mechanism for detecting psycho-physical state of a user based on the speech from the user; and

a spoken dialogue mechanism for carrying on a dialogue with said user based on the psycho-physical state of the user, detected by the psycho-physical detection mechanism from the speech from the user.

2. (Original) The system according to claim 1, wherein said spoken dialogue mechanism comprises:

a speech understanding mechanism for understanding the speech from the user based on the psycho-physical state of the user to generate a literal meaning of the speech; and

a voice response generation mechanism for generating a voice response to the user based on the literal meaning of the speech and the psycho-physical state of the user.

3. (Original) The system according to claim 2, wherein said speech understanding mechanism comprises:

at least one acoustic model for characterizing the acoustic properties of speech, each of said at least one acoustic model corresponding to some distinct characteristic related to a psycho-physical state of a speaker;

an acoustic model selection mechanism for selecting an acoustic model that is appropriate to according to the psycho-physical state detected by the psycho-physical state detection mechanism;

a speech recognizer for generating a transcription of spoken words recognized from the speech using the acoustic model selsected by the acoustic model selection mechanism; and

a language understanding mechanism for interpreting the literal meaning of the speech based on the transcription.

4. (Original) The system according to claim 2, wherein said voice response generation mechanism comprises:

a natural language response generator for generating a response based on an understanding of the transcription, said response being generated appropriately according to the psycho-physical state of the user;

a prosodic pattern determining mechanism for determining the prosodic pattern to be applied to said response that is considered as appropriate according to the psycho-physical state; and

3

a text-to-speech engine for synthesizing the voice response based on said response and said prosodic pattern.

5. (Original) The system according to claim 1, wherein said psycho-physical state detection mechanism comprises:

an acoustic feature extractor for extracting acoustic features from input speech data to generate at least one acoustic feature; and

a psycho-physical state classifier for classifying the input speech data into one or more psycho-physical states based on said at least one acoustic feature.

6. (Original) The system according to claim 5, further comprising:

at least one psycho-physical state model, each of said at least one psycho-physical state model corresponding to a single psycho-physical state and characterizing the acoustic properties of the single psycho-physical state; and

an off-line training mechanism for establishing said at least one psycho-physical model based on labeled training speech data.

7. (Original) The system according to claim 1, further comprising a dialogue manager that control the dialogue flow.

8. (Cancelled)

9. (Cancelled)

10. (Original) A method, comprising:

receiving, by a psycho-physical state detection mechanism, input speech data from a user;

detecting the psycho-physical state of the user from the input speech data;

understanding, by a speech understanding mechanism, the literal meaning of spoken words recognized from the input speech data based on the psycho-physical state of the user, detected by said detecting; and

generating, by a voice response generation mechanism, a voice response to the user based on the literal meaning of the input speech data and the psycho-physical state of the user.

4

11. (Original) The method according to claim 10, wherein said detecting comprises:

extracting, by a acoustic feature extractor, at least one acoustic feature from the input speech data; and

classifying, by a psycho-physical state classifier and based on said at least one feature, the input speech data into the psycho-physical state according to at least one psycho-physical state model.

12. (Original) The method according to claim 11, further comprising:

receiving, by an off-line training mechanism, labeled training data, wherein each of the data items in said labeled training data is labeled by a psycho-physical state; and

building said at least one psycho-physical state model using the labeled training data, each of the at least one psycho-physical state model corresponding to a single psycho-physical state and being established based on the data items in the labeled training data that have a label corresponding to the single psycho-physical state.

13. (Original) The method according to claim 10, wherein said understanding comprises:

selecting, by an acoustic model selection mechanism, an acoustic model, from at least one acoustic model, that is appropriate to according to the psycho-physical state, detected by said detecting, each of said at least one acoustic model corresponding to some distinct speech characteristic related to a psycho-physical state;

recognizing, by a speech recognizer, the spoken words from the input speech data using the acoustic model, selected by said selecting, to generate a transcription; and

interpreting, by a language understanding mechanism, the literal meaning of the spoken words based on the transcription.

14. (Original) The method according to claim 10, wherein said generating comprises:

constructing, by a natural language response generator, a natural language response based on an understanding of the transcription, said natural language response being constructed appropriately according to the psycho-physical state of the user;

determining, by a prosodic pattern determining mechanism, the prosodic pattern to be applied to said natural languiage response, wherein the prosodic pattern is considered to be appropriate according to the psycho-physical state; and

synthesizing, by a text-to-speech engine, the voice response based on said natural language response and said prosodic pattern.

5

15. (Cancelled)

16. (Cancelled)

17. (Cancelled)

18. (Original) A computer-readable medium encoded with a program, said program comprising:

receiving, by a psycho-physical state detection mechanism, input speech data from a user;

detecting the psycho-physical state of the user from the input speech data;

understanding, by a speech understanding mechanism, the literal meaning of spoken words recognized from the input speech data based on the psycho-physical state of the user, detected by said detecting; and

generating, by a voice response generation mecahnism, a voice response to the user based on the literal meaning of the input speech data and the psycho-physical state of the user.

19. (Original) The medium according to claim 18, wherein said detecting comprises:

extracting, by a acoustic feature extractor, at least one acoustic feature from the input speech data; and

classifying, by a psycho-physical state classifier and based on said at least one feature, the input speech data into the psycho-physical state according to at least one psycho-physical state model.

20. (Original) The medium according to claim 19, further comprising:

receiving, by an off-line training mechanism, labeled training data, wherein each of the data items in said labeled training data is labeled by a psycho-physical state; and

building said at least one psycho-physical state model using the labeled training data, each of the at least one psycho-physical state model corresponding to a single psycho-physical state and being established based on the data items in the labeled training data that have a label corresponding to the single psycho-physical state.

21. (Original) The medium according to claim 18, wherein said understanding comprises:

6

Guojun Zhou
Appl. No. 09/884,423

selecting, by an acoustic model selection mechanism, an acoustic model, from at least one acoustic model, that is appropriate to according to the psycho-physical state, detected by said detecting, each of said at least one acoustic model corresponding to some distinct speech characteristic related to a psycho-physical state;

recognizing, by a speech recognizer, the spoken words from the input speech data using the acoustic model, selected by said selecting, to generate a transcription; and.

interpreting, by a language understanding mechanism, the literal meaning of the spoken words based on the transcription.

22. (Original) The medium according to claim 18, wherein said generating comprises:

constructing, by a natural language response generator, a natural language response based on an understanding of the transcription, said natural language response being constructed appropriately according to the psycho-physical state of the user;

determining, by a prosodic pattern determining mechanism, the prosodic pattern to be applied to said natural languiage response, wherein the prosodic pattern is considered to be appropriate according to the psycho-physical state; and

synthesizing, by a text-to-speech engine, the voice response based on said natural language response and said prosodic pattern.

23. (Cancelled)

24. (Cancelled)